A Window to the World: Lessons Learned from NASA's Collaborative Metadata Curation Effort

Kaylin Bugbee

Katie Baynes, Valerie Dixon, Jeanne le Roux, Rahul Ramachandran, Dana Shum

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

14 December 2017

DSIG

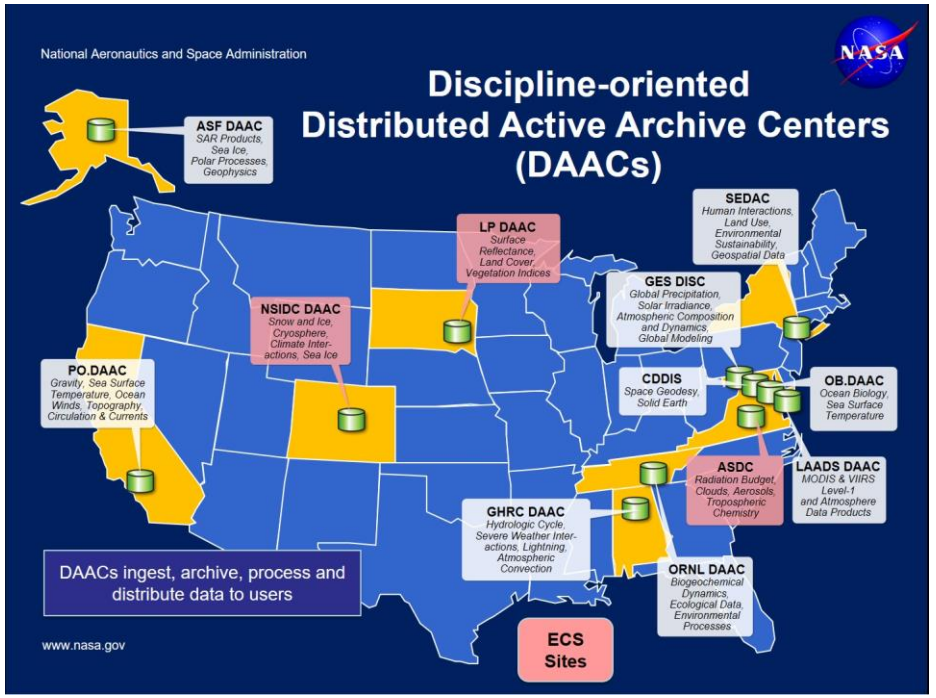
NASA Earth Science Data

NASA's Earth Observing System Data and Information System (EOSDIS)

Actively manages NASA's Earth science data

Includes:

- Science Investigator-led Processing Systems (SIPs)
- 12 distributed active archive centers (DAACs)



NASA Earth Science Data

Common Metadata Repository (CMR)

Authoritative management system for all EOSDIS metadata

Unified Metadata Model (UMM) crosswalks various dialects into CMR

Powers the Earthdata Search client

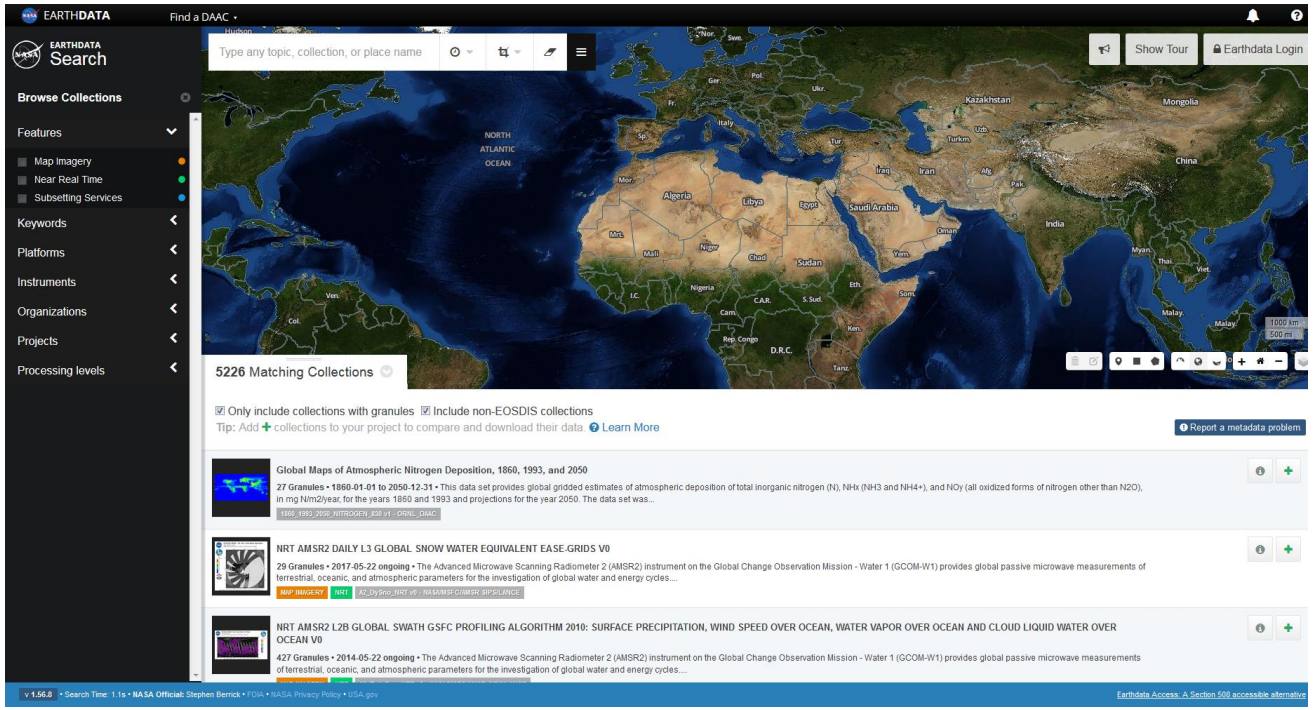
NASA Earth Science Metadata Curation Project

Goal

"Ensure data is documented precisely and fully, and access to data and services are useful and useable by a wide community of users."

Improve discoverability in CMR

Prepare for big data and bulk data download users



ESDIS mission statement goal #3 - https://earthdata.nasa.gov/earth-science-data-systems-program/about-the-esds-program

Metadata Curation Project: Who Are We?

Analysis and Review of CMR (ARC) team

Team is comprised of Earth Science data and metadata specialists

Collaborate extensively with

- DAAC Metadata Curators
- CMR Team
- Global Change Master Directory (GCMD)
- EOSDIS

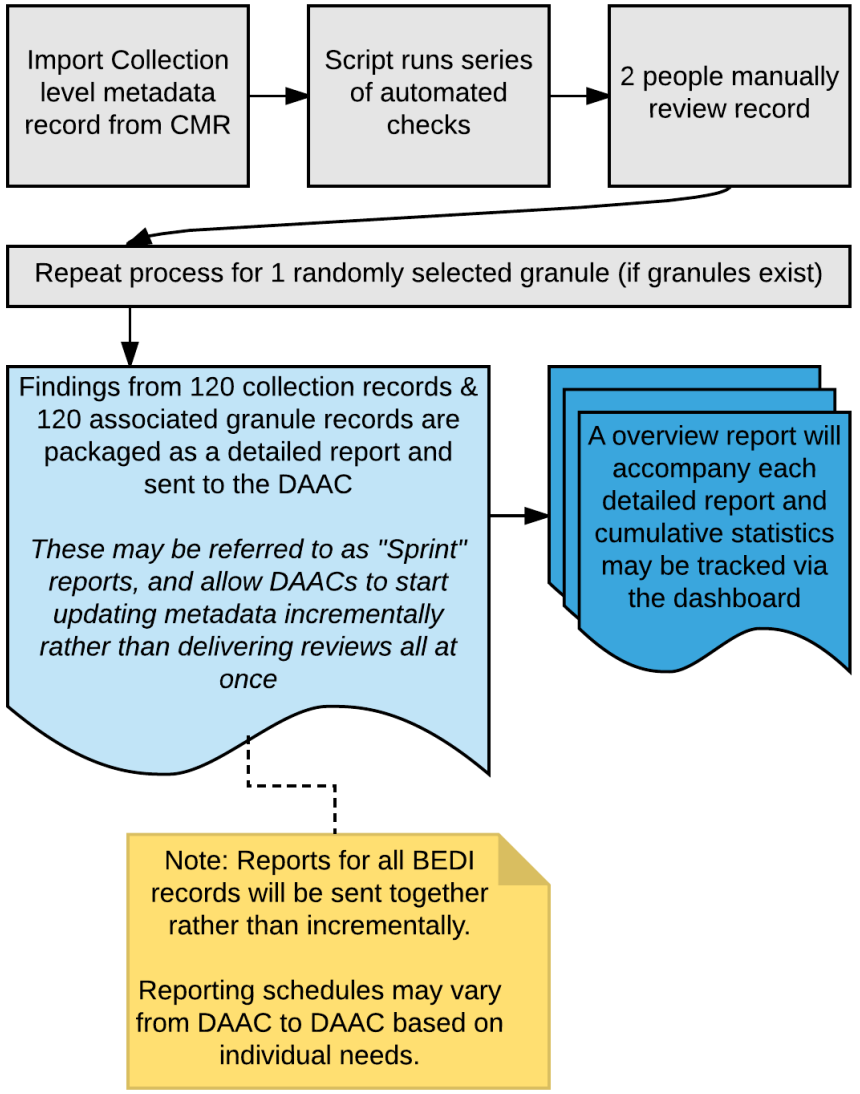
Metadata Curation Project: What We Do

ARC reviews collection and granule level metadata in the native format

Collaborate with DAAC metadata curators

Provide feedback to

- CMR team on UMM evolution
- GCMD team on keywords



Metadata Curation Project: Status

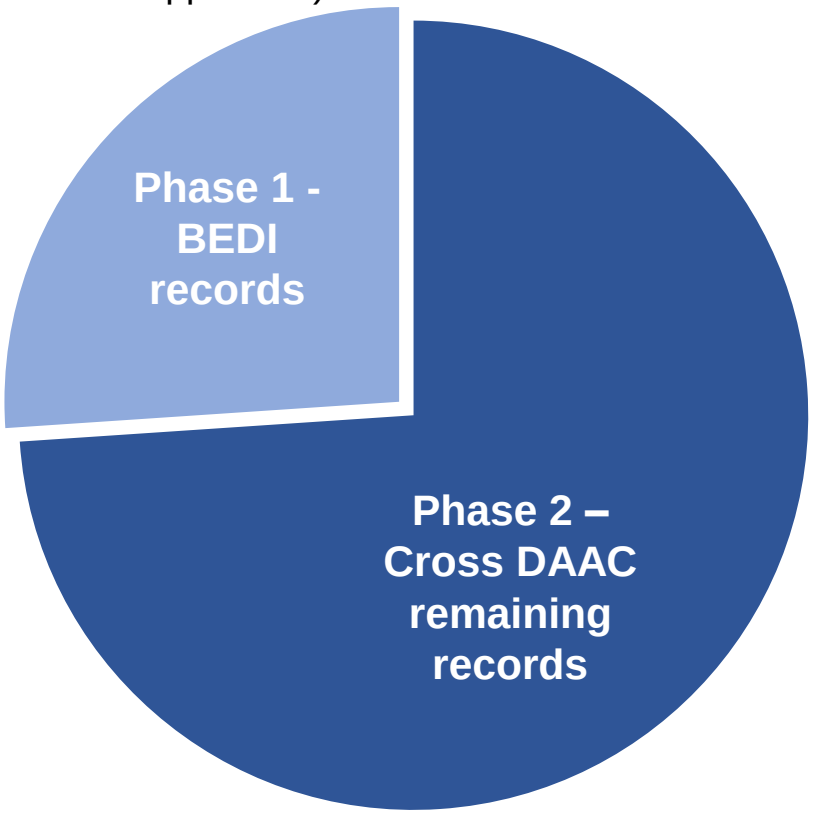
Reviewed over 25% of collection records to date

- 3,056 collection and granule records
- 4 dialects

Reviewed metadata records at all 12 DAACs

These reviews have provided a perspective on how pieces fit together and lessons learned about the broader enterprise

Over 1,800 collections reviewed in Phase 1 (plus random granules where applicable).



Approximately 5,100 collections to be reviewed in Phase 2.

Metadata Curation Project: Lessons Learned

Data and Services

Order tools

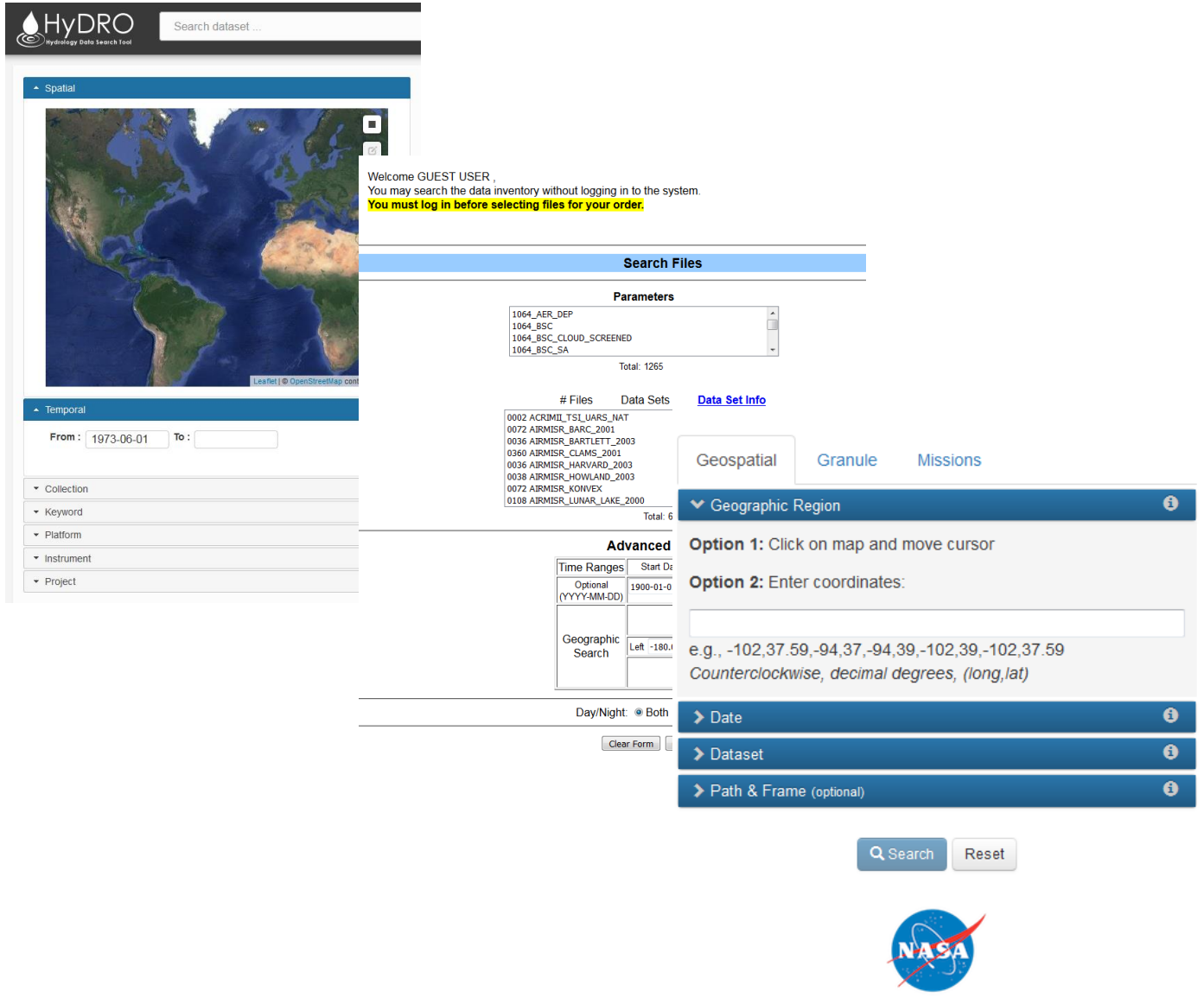
- Specialized tools for each DAAC

Granule level metadata

- Direct versus indirect access

File structure and variables

OPeNDAP



Metadata Curation Project: Lessons Learned

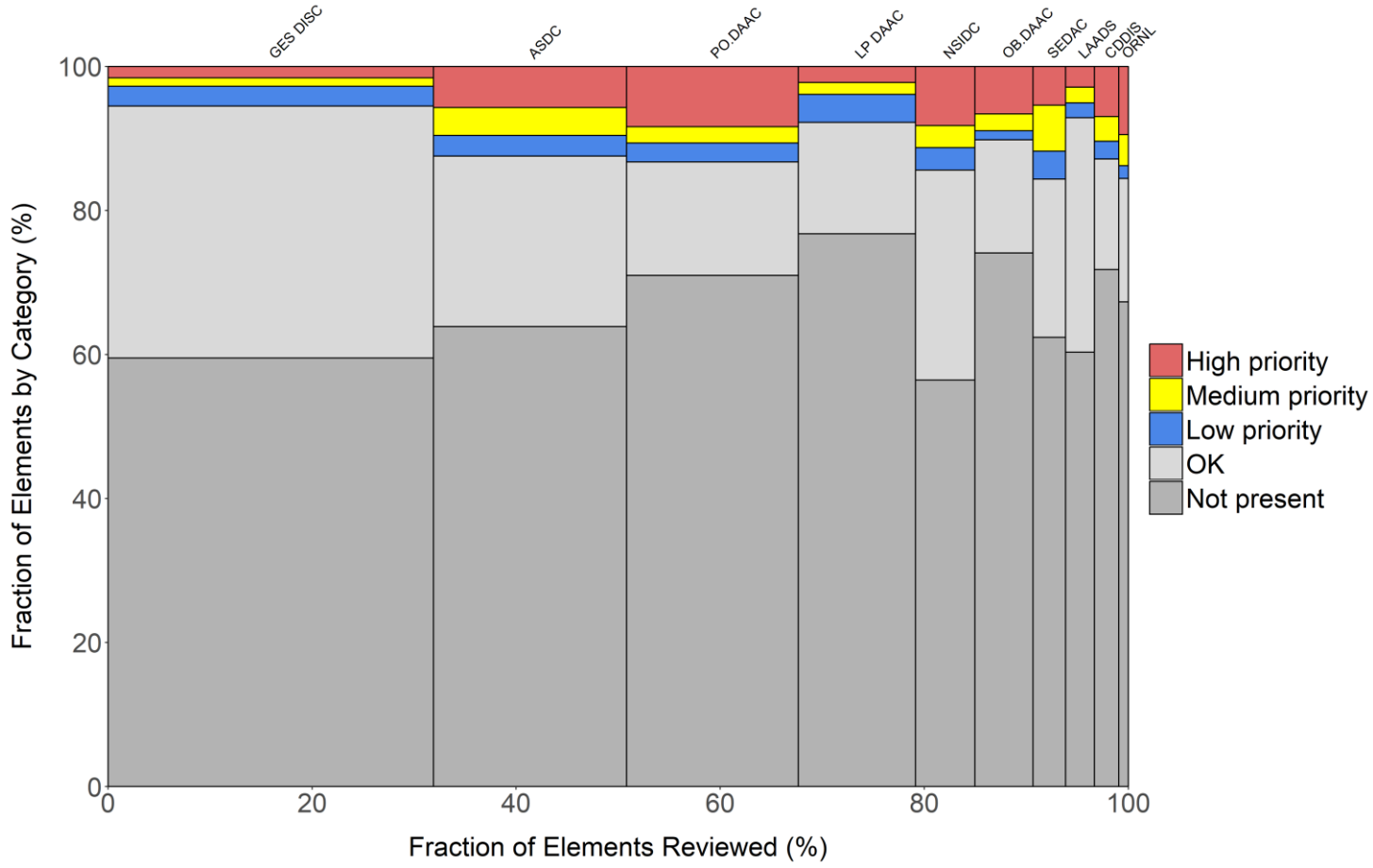
Metadata

Most metadata records only use around 40% of supported elements

Non satellite data in metadata

- Airborne data
- Field data

Science keyword updates



Metadata Curation Project: Lessons Learned

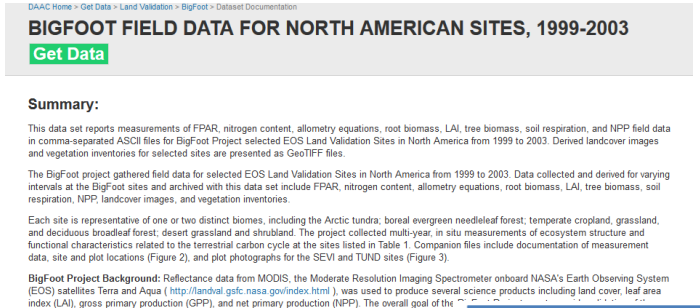
Documentation

Metadata documentation is provided in multiple places

- Prioritization on maintenance
- Guidance on which documents to use

Dataset documentation

- Different layouts
- Terminology



DAAC Home > Get Data > Land Validation > BigFoot > Dataset Documentation

BIGFOOT FIELD DATA FOR NORTH AMERICAN SITES, 1999-2003

Get Data

Summary:

This data set reports measurements of FPAR, nitrogen content, allometry equations, root biomass, LAI, tree biomass, soil respiration, and NPP field data in comma-separated ASCII files for BigFoot Project selected EOS Land Validation Sites in North America from 1999 to 2003. Derived landcover images and vegetation inventories for selected sites are presented as GeoTIFF files.

The BigFoot project gathered field data for selected EOS Land Validation Sites in North America from 1999 to 2003. Data collected and derived for varying intervals at the BigFoot sites and archived with this data set include FPAR, nitrogen content, allometry equations, root biomass, LAI, tree biomass, soil respiration, NPP, landcover images, and vegetation inventories.

Each site is representative of one or two distinct biomes, including the Arctic tundra; boreal evergreen needleleaf forest; temperate cropland, grassland, and deciduous broadleaf forest; desert grassland and shrubland. The project collected multi-year, in situ measurements of ecosystem structure and functional characteristics related to the terrestrial carbon cycle at the sites listed in Table 1. Companion files include documentation of measurement data, site and plot locations (Figure 2), and plot photographs for the SEVI and TUND sites (Figure 3).

BigFoot Project Background: Reflectance data from MODIS, the Moderate Resolution Imaging Spectrometer onboard NASA's Earth Observing System (EOS) satellites Terra and Aqua (http://landval.gsfc.nasa.gov/index.html), was used to produce several science products including land cover, leaf area index (LAI), gross primary production (GPP), and net primary production (NPP). The overall goal of the

products. To do this, BigFoot combined ground measurements, additional high-resolution remote-sens

tower sites representing different biomes to evaluate the effects of the spatial and temporal patterns of

BigFoot characterized up to a 7 x 7 km area (49 1-km MODIS pixels) surrounding the CO_2 flux towers

sampling design allowed the Project to examine scales and spatial patterns of these properties, the ir

products, and provided for a field-based ecological characterization of the flux tower footprint. BigFoot

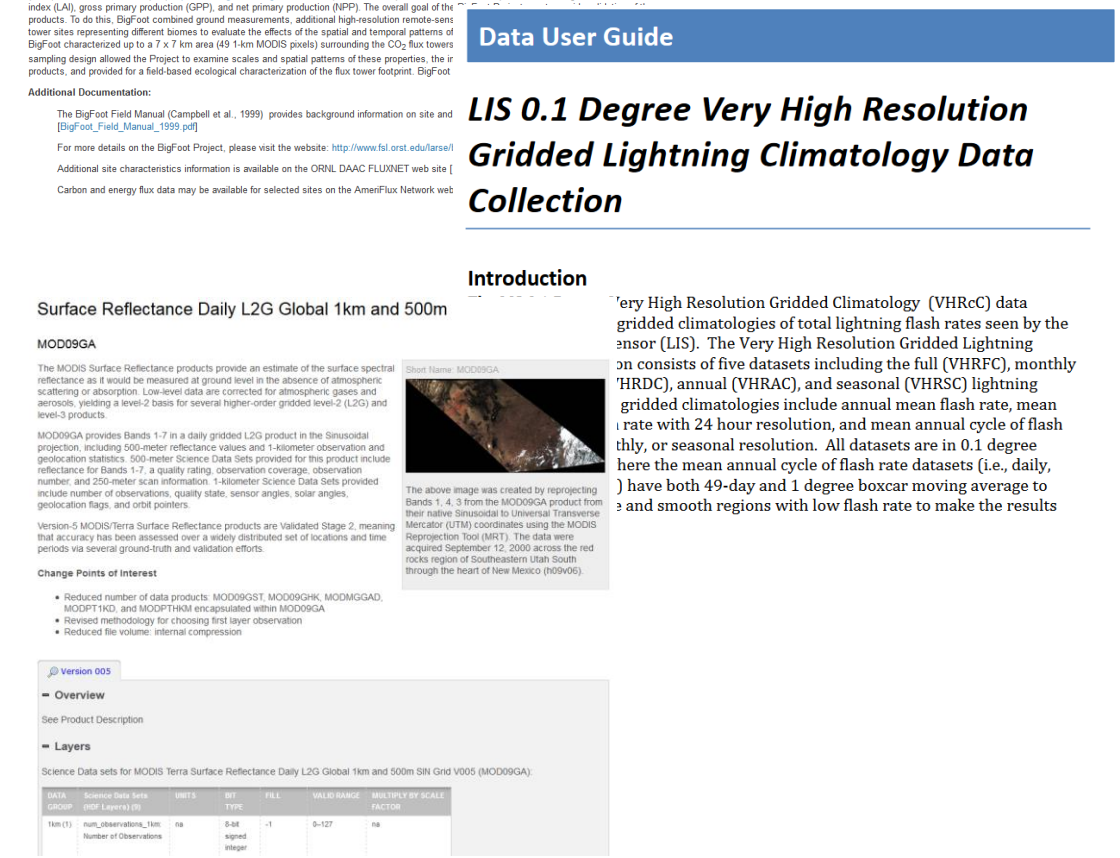
Additional Documentation:

The BigFoot Field Manual (Campbell et al., 1999) provides background information on site and [BigFoot_Field_Manual_1999.pdf]

For more details on the BigFoot Project, please visit the website: http://www.fsl.orst.edu/larse/l

Additional site characteristics information is available on the ORNL DAAC FLUXNET web site [

Carbon and energy flux data may be available for selected sites on the AmeriFlux Network web



Data User Guide

LIS 0.1 Degree Very High Resolution Gridded Lightning Climatology Data Collection

Introduction

ery High Resolution Gridded Climatology (VHRcC) data
gridded climatologies of total lightning flash rates seen by the
ensor (LIS). The Very High Resolution Gridded Lightning
on consists of five datasets including the full (VHRFC), monthly
HRDC), annual (VHRAC), and seasonal (VHRSC) lightning
gridded climatologies include annual mean flash rate, mean
rate with 24 hour resolution, and mean annual cycle of flash
thly, or seasonal resolution. All datasets are in 0.1 degree
here the mean annual cycle of flash rate datasets (i.e., daily,
) have both 49-day and 1 degree boxcar moving average to
e and smooth regions with low flash rate to make the results

Surface Reflectance Daily L2G Global 1km and 500m

MOD09GA

The MODIS Surface Reflectance products provide an estimate of the surface spectral reflectance as it would be measured at ground level in the absence of atmospheric scattering or absorption. Low-level data are corrected for atmospheric gases and aerosols, yielding a level-2 basis for several higher-order gridded level-2 (L2G) and level-3 products.

MOD09GA provides Bands 1-7 in a daily gridded L2G product in the Sinusoidal projection, including 500-meter reflectance values and 1-kilometer observation and geolocation statistics. 500-meter Science Data Sets provided for this product include reflectance for Bands 1-7, a quality rating, observation coverage, observation number, and 250-meter scan information. 1-kilometer Science Data Sets provided include number of observations, quality state, sensor angles, solar angles, geolocation flags, and orbit pointers.

Version-5 MODIS/Terra Surface Reflectance products are Validated Stage 2, meaning that accuracy has been assessed over a widely distributed set of locations and time periods via several ground-truth and validation efforts.

Short Name: MOD09GA

The above image was created by reprojecting Bands 1, 4, 3 from the MOD09GA product from their native Sinusoidal to Universal Transverse Mercator (UTM) coordinates using the MODIS Reprojection Tool (MRT). The data were acquired September 12, 2000 across the red rocks region of Southeastern Utah South through the heart of New Mexico (h09v06).

Change Points of Interest

- Reduced number of data products: MOD09GST, MOD09GHK, MODMGGAD, MODPT1KD, and MODPTHKM encapsulated within MOD09GA
- Revised methodology for choosing first layer observation
- Reduced file volume: internal compression

Version 005

Overview

See Product Description

Layers

Science Data sets for MODIS Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V005 (MOD09GA):

DATA GROUP	Science Data Sets (HDF Layers) (9)	UNITS	BIT TYPE	FILL	VALID RANGE	MULTIPLY BY SCALE FACTOR
1km (1)	num_observations_1km: Number of Observations	na	8-bit signed integer	-1	0–127	na

Metadata Curation Project: Lessons Learned

Policies

Metadata authors and data center curators are looking for clear, concise policies to make generating metadata and other data center tasks easier

Guidance is available

Provided in a lot of different places at various levels of detail

Metadata Curation Project: Moving Forward

What Are the Next Steps?

Metadata

- Collaborating across all stakeholders on consolidating documentation, refining policies
  - Created a metadata curation mailing list and a curation channel on Slack to keep the conversation going
- Refining bulk update capabilities for common metadata fixes
- Developing and refining curation tools to help data curators and owners validate, improve, and correct metadata
- Further evolution of metadata documentation (UMM) and schema implementation
  - Developing recommendations for non-satellite metadata

Questions?

Contact Information:

Kaylin Bugbee
Kaylin.m.Bugbee@nasa.gov